An Overview of Effect Sizes and Meta-analysis

*KTDRR Research Evidence Training*

Ryan Williams, PhD, Principal Researcher
Joshua R. Polanin, PhD, Principal Researcher
American Institutes for Research
Introduction

▪ Ryan Williams
  • Principal Researcher, American Institutes for Research (AIR)
  • Principal Investigator: Institute of Education Sciences (IES)-funded meta-analysis focusing on heterogeneity in mathematics intervention effects
  • Co-Principal Investigator: IES-funded training grant on meta-analysis

▪ Josh Polanin
  • Principal Researcher, AIR
  • Principal Investigator: two National Institute of Justice (NIJ)-funded reviews on school violence
  • Co-Principal Investigator: IES training grant on meta-analysis; IES-funded review on college aid
  • Project Director: What Works Clearinghouse Statistics, Website, and Training (SWAT) contract
Overview

- **Effect Size**
  - What is an effect size, and why is it important in meta-analysis?
  - How to calculate an effect size depending on what the underlying data are
  - Three varieties: standardized mean-difference, odds ratio, correlation

- **Meta-analysis**
  - Why synthesize effect sizes?
  - Basics of fixed- and random-effects models
  - Assessing heterogeneity among effect sizes
  - Explaining heterogeneity based on units, treatments, outcomes, and settings (UTOS)

- If time allows—questions at the end
What is an effect size, and why is it important in meta-analysis?

- Start with a thought experiment: Imagine you are interested in understanding whether various programs—all using the same basic tenets—have an impact on an outcome domain.
  - “The purpose of this review was to determine whether therapeutic riding and hippotherapy improve *balance*…” (among many other outcomes)

- How *balance* is measured across each included study probably varies—but each measure focuses on the same underlying construct.
What is an effect size, and why is it important in meta-analysis? (Cont.)

- Meta-analysis expresses the results of each study using a quantitative index of effect size.
- Effect sizes are measures of the strength or magnitude of a relationship of interest.
- Effect sizes have the advantage of being comparable (i.e., they estimate the same thing) across all the studies and therefore can be summarized across studies in the meta-analysis.
What effect size should be calculated?

- Effect sizes can be expressed in many different metrics.
  - $d$, $r$, $OR$, $RR$, etc.
    - The decision about which metric to use is based on what the primary authors report.

- Effect sizes can be unstandardized or standardized.
  - Unstandardized = expressed in measurement units (do not have the properties we need!)
  - Standardized = expressed in standardized measurement units

- Effect sizes should be accompanied by their standard errors; these get calculated along with the effect size and will be used in meta-analytic calculations.
The $d$ family

- The standardized mean difference.
- Used when we are interested in two-group comparisons using means.
- Groups could be two experimental groups or, in an observational study, two groups of interest, such as boys versus girls.
The \( d \) family (Cont.)

- Notation:

  Group means: \( \bar{X}_{G1}, \bar{X}_{G2} \)
  Group sample sizes: \( n_{G1}, n_{G2} \)
  Total sample size: \( N = n_{G1} + n_{G2} \)
  Group standard deviations: \( s_{G1}, s_{G2} \)
The $d$ family (Cont.)

- Notation:

  Group means: $\bar{X}_{G1}, \bar{X}_{G2}$
  Group sample sizes: $n_{G1}, n_{G2}$
  Total sample size: $N = n_{G1} + n_{G2}$
  Group standard deviations: $s_{G1}, s_{G2}$

$$ES_{sm} = \frac{\bar{X}_{G1} - \bar{X}_{G2}}{s_p}$$

Pooled sample standard deviation (i.e., a weighted average standard deviation)
The \( d \) family (Cont.)

- The regular effect size formula is biased, especially with small samples.
- We can easily apply a correction for this bias:

\[
ES'_{sm} = \left[ 1 - \frac{3}{4N - 9} \right] ES_{sm}
\]

- This produces a Hedges’ \( g \) effect size, or bias-corrected estimate.
The $d$ family (Cont.)

- Each effect size needs a measure of precision:

$$SE_{sm} = \sqrt{\frac{n_{G1} + n_{G2}}{n_{G1}n_{G2}} + \frac{ES'_{sm}^2}{2(n_{G1} + n_{G2})}}$$
The \( d \) family (Cont.)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Multisystemic therapy</th>
<th>Usual services</th>
<th>( F )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td>Ultimate outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arrests*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M )</td>
<td></td>
<td>0.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( SD )</td>
<td></td>
<td>1.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incarceration* (in weeks)</td>
<td></td>
<td>5.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( M )</td>
<td></td>
<td>13.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( SD )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRD</td>
<td></td>
<td>11.5</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>( M )</td>
<td></td>
<td>15.7</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>( SD )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The $d$ family (Cont.)

- We can take the values reported in Table 1 and compute an effect size and its standard error:

\[
sp = \sqrt{\frac{(43 - 1)13.9^2 + (41 - 1)19.1^2}{(43 - 1) + (41 - 1)}} = 16.6
\]

\[
es_{sm} = \frac{5.8 - 16.2}{16.6} = -.63
\]

\[
es'_{sm} = \left[1 - \frac{3}{4(43 + 41) - 9}\right] - .63 = -.62
\]

\[
se_{sm} = \sqrt{\frac{43 + 41}{43 \times 41} + \frac{-.62^2}{2(43 + 41)}} = .22
\]
The $r$ family

- The correlation coefficient, or $r$ family effects, may be appropriate when …
  - studies have a continuous outcome measure,
  - study designs assess the relation between a quantitative predictor and the outcome (possibly controlling for covariates), or
  - the analysis uses regression (or the general linear model).
The $r$ family (Cont.)

- When using the correlation, you (or a computer program) will do these two things:
  1. Translate it into Fisher’s $z$:

$$Z_r = \frac{1}{2} \ln \left( \frac{1 + r}{1 - r} \right)$$

  2. Calculate the standard error of $Z$:

$$SE_z = \frac{1}{\sqrt{n - 3}}$$
The $r$ family (Cont.)

- Practically—the formulas are important to know, but you probably won’t use them directly.

- Instead—if the correlation is of interest, you will need to locate it in the study and locate its sample size.
The odds ratio family(ish)

- Consider a study in which a treatment group (Tx) and a control group (Cx) are compared with respect to the frequency of a binary characteristic among the participants.

- In each group, we will count how many participants satisfy the binary outcome of interest (e.g., passing a test, graduating, being cured of a disease, etc.).

- The odds ratio is one of a few effect sizes that can be calculated in these scenarios (but it is not the only one).
The odds ratio family(ish) (Cont.)

<table>
<thead>
<tr>
<th>Study Group</th>
<th>Success</th>
<th>Failure</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>5</td>
<td>14</td>
<td>19</td>
</tr>
<tr>
<td>Comparison</td>
<td>6</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>TOTAL</td>
<td>11</td>
<td>26</td>
<td>37</td>
</tr>
</tbody>
</table>

Odds of treatment “success” \( \frac{5}{14} = .36 \)

Odds of comparison “success” \( \frac{6}{12} = .50 \)

The “odds ratio” is literally just that—a ratio of two odds—in this case, the odds of treatment success divided by the odds of comparison group success.

\[ \frac{.36}{.50} = .72 \]

The odds of success in treatment are .72 times the odds of success in control.
The odds ratio family(ish) (Cont.)

- When using the correlation, you (or a computer program) will do these two things:
  1. Transform it to the log odds ratio:
     \[ \log \text{odds ratio} = \ln(OR) \]
  2. Calculate the log odds ratio standard error:
     \[
     SE_{\ln(OR)} = \sqrt{\frac{1}{Cell \ "a"} + \frac{1}{Cell \ "b"} + \frac{1}{Cell \ "c"} + \frac{1}{Cell \ "d"}}
     \]

<table>
<thead>
<tr>
<th></th>
<th>S</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tr</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Co</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>
Meta-analysis Introduction

- Computing an average effect is done in two general ways:
  - Fixed-effects model
  - Random-effects model
- Both approaches typically will use an estimate of precision to calculate a weighted mean effect size.
- BUT, the two approaches differ in how they characterize those weights.
Meta-analysis Introduction (Cont.)

- Computing an average effect is done in two general ways:
  - Fixed-effects model
  - Random-effects model
- Both approaches typically will use an estimate of precision to calculate a weighted mean effect size.
- BUT, the two approaches differ in how they characterize those weights.
Meta-analysis Slides

- The fixed-effects model considers one of variation: sampling variance.

\[
\overline{ES_{FE}} = \frac{\sum_{i=1}^{j} (w_i ES_i)}{\sum_{i=1}^{j} (w_i)} \\
SE_{FE} = \sqrt{\frac{1}{\sum_{i=1}^{j} w_i}} \\
\]

\[
w_i = \frac{1}{SE_i^2} \\
\]
Meta-analysis Slides (Cont.)

- This model is helpful when the effect sizes are homogeneous.
- That is, the only reason they are identical is because they have different samples.
- This is a strong assumption.
The random-effects model considers two sources of variation: sampling variance and between-study variance.

\[ \bar{ES}_{RE} = \frac{\sum_{i=1}^{j} (w_i ES_i)}{\sum_{i=1}^{j} (w_i)} \]

\[ SE_{RE} = \sqrt{\frac{1}{\sum_{i=1}^{j} w_i}} \]

\[ w_i = \frac{1}{SE_i^2 + \tau^2} \]
The random-effects model considers **two** sources of variation: sampling variance and between-study variance.

\[ \bar{ES}_{RE} = \frac{\sum_{i=1}^{j} (w_i ES_i)}{\sum_{i=1}^{j} (w_i)} \]

\[ SE_{RE} = \sqrt{\frac{1}{\sum_{i=1}^{j} w_i}} \]

\[ w_i = \frac{1}{SE_i^2 + \hat{\tau}^2} \]

Estimated heterogeneity
Because the random-effects model uses two sources of variation, the standard error around the mean effect will be larger than the standard error for an equally sized fixed-effects model.

BUT, unless the multiple effect sizes are coming from maximum-control conditions (i.e., carefully scripted, laboratory-type replications), it is hard to rule out a random-effects model.

For this reason, we strongly encourage using a random-effects model. If your data are statistically homogeneous, it won’t hurt you!
What is effect size heterogeneity?

Source: Hennessy & Tanner-Smith, *Prevention Science*, 2015
Quantifying heterogeneity

- Three primary statistics:
  - $Q$ – Tells us if the variation is different from chance.
  - $\tau^2$ – Tells us the magnitude of the variation.
  - $I^2$ – Tells us the proportion of true variation among effects (taking into account the possibility of random variation).

- These allow us to turn visual information into quantitative information.
Explaining heterogeneity

- What to do with sufficient heterogeneity?
- One-way moderator analyses (old approach)
- One-variable meta-regression (old approach)
- Multiple variable meta-regression (best approach)
  - Control for confounding factors
  - Reduce type 1 errors
  - Easier to interpret
Thank you!

Please take a few minutes to respond to the brief Evaluation Survey:


Ryan Williams: rwilliams@air.org
Joshua Polanin: jpolanin@air.org